Hebbian learning of recurrent connections: a geometrical perspective

Mathieu N. Galtier [*] Olivier D. Faugeras [†]
Paul C. Bressloff [‡]

January 10, 2013

Abstract: We show how a Hopfield network with modifiable recurrent connections undergoing slow Hebbian learning can extract the underlying geometry of an input space. First, we use a slow/fast analysis to derive an averaged system whose dynamics derives from an energy function and therefore always converges to equilibrium points. The equilibria reflect the correlation structure of the inputs, a global object extracted through local recurrent interactions only. Second, we use numerical methods to illustrate how learning extracts the hidden geometrical structure of the inputs. Indeed, multidimensional scaling methods make it possible to project the final connectivity matrix on to a distance matrix in a high-dimensional space, with the neurons labelled by spatial position within this space. The resulting network structure turns out to be roughly convolutional. The residual of the projection defines the non-convolutional part of the connectivity which is minimized in the process. Finally, we show how restricting the dimension of the space where the neurons live gives rise to patterns similar to cortical maps. We motivate this using an energy efficiency argument based on wire length minimization. Finally, we show how this approach leads to the emergence of ocular

[*] Corresponding author: mathieu.galtier@inria.fr. NeuroMathComp Project Team, INRIA Sophia-Antipolis Méditerranée, 2004 route des Lucioles-BP 93, 06902 Sophia Antipolis, France

[†] NeuroMathComp Project Team, INRIA Sophia-Antipolis Méditerranée, 2004 route des Lucioles-BP 93, 06902 Sophia Antipolis, France

[‡] Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, Utah 84112, USA. Mathematical Institute, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, UK

dominance or orientation columns in primary visual cortex. In addition, we establish that the non-convolutional (or long-range) connectivity is patchy, and is co-aligned in the case of orientation learning.

Keywords: correlation-based Hebbian learning, Hopfield networks, temporal averaging, energy minimization, multidimensional scaling, cortical maps

1 Introduction

Activity-dependent synaptic plasticity is generally thought to be the basic cellular substrate underlying learning and memory in the brain. Donald Hebb [Hebb, 1949] postulated that learning is based on the correlated activity of synaptically connected neurons: if both neurons A and B are active at the same time, then the synapses from A to B and B to A should be strengthened proportionally to the product of the activity of A and B. However, as it stands, Hebb's learning rule diverges. Therefore, various modification of Hebb's rule have been developed, which basically take one of three forms (see [Gerstner and Kistler, 2002] and [Dayan and Abbott, 2001]): first, a decay term can be added to the learning rule so that each synaptic weight is able to "forget" what it previously learned. Second, each synaptic modification can be normalized or projected on different subspaces. These constraint–based rules may be interpreted as implementing some form of competition for energy between dendrites and axons, see [Miller, 1996, Miller and MacKay, 1996] and [Ooyen, 2001] for details. Third, a sliding threshold mechanism can be added to Hebbian learning. For instance, a post-synaptic threshold rule consists in multiplying the presynaptic activity and the subtraction of the average postsynaptic activity from its current value, which is referred as covariance learning ([Sejnowski and Tesauro, 1989]). Probably the best known of these rules is the BCM rule [Bienenstock et al., 1982]. It should be noted that history-based rules can also be defined without changing the qualitative dynamics of the system: instead of considering the instantaneous value of the neurons' activity, these rules consider its weighted mean over a time window (see [Földiák, 1991, Wallis and Baddeley, 1997]). Recent experimental evidence suggests that learning may also depend upon the precise timing of action potentials [Bi and Poo, 2001]. Contrary to most Hebbian rules that only detect correlations, these rules can also encode causal relationships in the patterns of neural activation. However, the mathematical treatment of these spike timing dependent rules is much more difficult than rate based ones.

Hebbian-like learning rules have often been studied within the framework of unsupervised feedfoward neural networks [Oja, 1982, Bienenstock et al., 1982, Miller and MacKay, 1996, Dayan and Abbott, 2001]. They also form the basis of most weight-based models of cortical development, assuming fixed lateral connectivity (e.g. mexican hat) and modifiable vertical connections (see the review of [Swindale, 1996])[1]. In these developmental models, the statistical structure of input correlations provides a mechanism for spontaneously breaking some underlying symmetry of the neuronal receptive fields leading to the emergence of feature selectivity. When such correlations are combined with fixed intracortical interactions, there is a simultaneous breaking of translation symmetry across cortex leading to the formation of a spatially periodic cortical feature map. A related mathematical formulation of cortical map formation has been developed in [Takeuchi and Amari, 1979, Bressloff, 2005] using the theory of self–organizing neural fields. Although very irregular, the two-dimensional cortical maps observed at a given stage of development, can be unfolded in higher dimensions to get smoother geometrical structures. Indeed, [Bressloff et al., 2001] suggested that the network of orientation pinwheels in V1 is a direct product between a circle for orientation preference and a plane for position, based on a modification of the icecube model of Hubel and Wiesel [Hubel and Wiesel, 1977]. From a more abstract geometrical perspective, Petitot [Petitot, 2003] has associated such a structure to a 1-jet space and used this to develop some applications to computer vision. More recently, [Bressloff and Cowan, 2003] and [Chossat and Faugeras, 2009] have considered more complex geometrical structures such as spheres and hyperbolic surfaces that incorporate additional stimulus features such as spatial frequency and textures, respectively.

In this paper, we show how geometrical structures related to the distribution of inputs can emerge through unsupervised Hebbian learning applied to recurrent connections in a rate-based Hopfield network. Throughout this paper, the inputs are presented as an external non-autonomous forcing to the system and not an initial condition as is often the case in Hopfield networks. It has previously been shown that, in the case of a single fixed input, there exists an energy function that describes the joint gradient dynamics of the activity and weight variables [Dong and Hopfield, 1992]. This implies that the system converges to an equilibrium during learning. We use averaging theory to generalize the above result to the case of multiple inputs, under the adiabatic assumption that Hebbian learning

[1] There have only been a few computational studies that consider the joint development of lateral and vertical connections [Bartsch and Van Hemmen, 2001, Miikkulainen et al., 2005].

occurs on a much slower time scale than both the activity dynamics and the sampling of the input distribution. We then show that the equilibrium distribution of weights, when embedded into $\mathbb{R}^k$ for sufficiently large integer k, encodes the geometrical structure of the inputs. Finally, we numerically show that the embedding of the weights in two dimensions ($k = 2$) gives rise to patterns that are qualitatively similar to experimentally observed cortical maps, with the emergence of features columns and patchy connectivity. Although the mathematical formalism we introduce here could be extended to most of the rate-based Hebbian rules in the literature, we present the theory for Hebbian learning with decay because of the simplicity of the resulting dynamics.

Note that the use of geometrical objects to describe the emergence of connectivity patterns has previously been put forward by Amari in a different context. Based on the theory of information geometry, Amari considers the geometry of the set of all the networks and defines learning as a trajectory on this manifold for perceptron networks in the framework of supervised learning [Amari, 1998] or for unsupervised Boltzmann Machines [Amari et al., 1992]. He uses differential and Riemannian geometry to describe an object which is at a larger scale than the cortical maps this paper is focusing on.

Moreover, Zucker and colleagues are currently developing a non-linear dimensionality reduction approach to caracterize the statistics of natural visual stimuli (see [Lawlor and Zucker, 2010, Coifman et al., 2005]). Although they do not use learning neural networks and stay closer to the field of computer vision than this paper, it turns out their approach is similar to the geometrical embedding approach we are using.

The structure of the paper is as follows. In section 2, we formally introduce the model. We derive the averaged system in section 3, which then allows us to study the stability of the learning dynamics in the presence of multiple inputs by constructing an appropriate energy function. We adress stability in section 4. In section 5 we determine the geometrical structure of the equilibrium weight distribution and show how it reflects the structure of the inputs. We also relate this approach to the emergence of cortical maps. Finally, the results are discussed in section 6.

2 Model

2.1 Neural network evolution

A neural mass corresponds to a mesoscopic coherent group of neurons. It is convenient to consider them as building blocks, first for computational simplicity, second for their direct relationship to macroscopic measurements of the brain (EEG, MEG and Optical imaging) which average over numerous neurons, and third because one can functionally define coherent groups of neurons within cortical columns. For each neural mass $i \in \{1..N\}$, define the mean membrane potential $V_i(t)$ at time t. The instantaneous population firing rate $\nu_i(t)$ is linked to the membrane potential through the relation $\nu_i(t) = s\big(V_i(t)\big)$, where s is a smooth sigmoid function. In the following, we choose

$$s(v) = \frac{S_m}{1 + \exp\big(-4S_m'(v - \phi)\big)}, \tag{1}$$

where S_m, S_m' and ϕ are respectively the maximal firing rate, the maximal slope and the offset of the sigmoid.

Consider a Hopfield network of neural masses described by the equation

$$\frac{dV_i}{dt}(t) = -\alpha V_i(t) + \sum_{j=1}^{N} W_{ij}(t)\, s\big(V_j(t)\big) + I_i(t). \tag{2}$$

The first term roughly corresponds to the intrinsic dynamics of the neural mass: it decays exponentially to zero at a rate α if it receives neither external inputs nor spikes from the other neural masses. We will fix the units of time by setting $\alpha = 1$. The second term corresponds to the rest of the network sending information through spikes to the given neural mass i, with $W_{ij}(t)$ the effective synaptic weight from neural mass j. The synaptic weights are time–dependent because they evolve according to a continuous time Hebbian learning rule (see below). The third term $I_i(t)$ corresponds to an external input to neural mass i, e.g. information extracted by the retina or thalamo-cortical connections. We take the inputs to be piecewise constant in time, that is, at regular time intervals a new input is presented to the network. In this paper, we will assume that the inputs are chosen by peridodically cycling through a given set of M inputs. An alternative approach would be to randomly select each input from a given probability distribution [Geman, 1979]. It is convenient to introduce vector notation by representing the time–dependent membrane potentials by $V \in C^1(\mathbb{R}_+, \mathbb{R}^N)$, the time–dependent external inputs by $I \in C^0(\mathbb{R}_+, \mathbb{R}^N)$, and the time–dependent network weight matrix by $W \in C^1(\mathbb{R}_+, \mathbb{R}^{N\times N})$. We can then rewrite the above system of ordinary differential equations as a single vector-valued equation

$$\frac{dV}{dt} = -V + W \cdot S(V) + I, \tag{3}$$

where $S : \mathbb{R}^N \to \mathbb{R}^N$ corrresponds to the term by term application of the sigmoid S, i.e. $S(V)_i = s(V_i)$.

2.2 Correlation-based Hebbian learning

The synaptic weights are assumed to evolve according to a correlation–based Hebbian learning rule of the form

$$\frac{dW_{ij}}{dt} = \epsilon(s(V_i)s(V_j) - \mu W_{ij}), \tag{4}$$

where ϵ is the learning rate, and we have included a decay term in order to stop the weights from diverging. In order to rewrite the above equation in a more compact vector form, we introduce the tensor (or Kronecker) product $S(V) \otimes S(V)$ so that in component form

$$[S(V) \otimes S(V)]_{ij} = S(V)_i S(V)_j, \tag{5}$$

where S is treated as a mapping from $\mathbb{R}^N$ to $\mathbb{R}^N$. The tensor product implements Hebb's rule that synaptic modifications involve the product of postynaptic and presynaptic firing rates. We can then rewrite the combined voltage and weight dynamics as the following non–autonomous (due to time–dependent inputs) dynamical system:

$$\Sigma : \quad \left\{ \begin{array}{rcl} \dfrac{dV}{dt} & = & -V + W \cdot S(V) + I \\[2ex] \dfrac{dW}{dt} & = & \epsilon\Big(S(V) \otimes S(V) - \mu W\Big). \end{array} \right. \tag{6}$$

Let us make few remarks about the existence and uniqueness of solutions. First, boundedness of S implies boundedness of the system Σ. More precisely, if I is bounded, then the solutions are bounded. To prove this, note that the right hand side of the equation for W is the sum of a bounded term and a linear decay term in W. Therefore, W is bounded and hence the term $W \cdot S(V)$ is also bounded. The same reasoning applies to V. S being Lipschitz continuous implies that the right hand side of the system is Lipschitz. This is sufficient to prove existence and uniqueness of the solution by applying the Cauchy-Lipschitz theorem. In the following, we will derive an averaged autonomous dynamical system Σ', which will be well-defined for the same reasons.

3 Averaging the system

We will show that system Σ can be approximated by an autonomous Cauchy problem which will be much more convenient to handle. This averaging method makes the most of multiple time–scales in the system. First, it is natural to consider that learning occurs on a much slower time–scale than the evolution of the membrane potentials (as determined by α), i.e.

$$\epsilon \ll 1. \tag{7}$$

Second, an additional time-scale arises from the rate at which the inputs are sampled by the network. That is, the network cycles periodically through M fixed inputs, with the period of cycling given by T. It follows that I is T–periodic, piecewise constant. We assume that the sampling rate is also much slower than the evolution of the membrane potentials,

$$\frac{M}{T} \ll 1. \tag{8}$$

Finally, we assume that the period T is small compared to the time-scale of the learning dynamics,

$$\epsilon \ll \frac{1}{T}. \tag{9}$$

We can now simplify the system Σ by applying Tikhonov's theorem for slow/fast systems, and then classical averaging methods for periodic systems.

3.1 Tikhonov's theorem

Tikhonov's theorem ([Tikhonov, 1952] and [Verhulst, 2007] for a clear introduction) deals with slow/fast systems. It says the following:

Theorem 3.1. *Consider the initial value problem*

$$\begin{array}{l} \dot{x} = f(x,y,t),\ x(0) = x_0,\ x \in \mathbb{R}^n, t\ \in \mathbb{R}_+ \\ \epsilon\dot{y} = g(x,y,t),\ y(0) = y_0,\ y \in \mathbb{R}^m \end{array}$$

Assume that:

1. *A unique solution of the initial value problem exists and we suppose, this holds also for the reduced problem*

$$\begin{array}{l} \dot{x} = f(x,y,t),\ x(0) = x_0 \\ 0 = g(x,y,t) \end{array}$$

   with solutions $\bar{x}(t)$, $\bar{y}(t)$.

2. *The equation $0 = g(x,y,t)$ is solved by $\bar{y}(t) = \phi(x,t)$, where $\phi(x,t)$ is a continuous function and an isolated root. Also suppose that $\bar{y}(t) = \phi(x,t)$ is an asymptotically stable solution of the equation $\frac{dy}{d\tau} = g(x,y,\tau)$ that is uniform in the parameters $x \in \mathbb{R}^n$ and $t \in \mathbb{R}_+$.*

3. *$y(0)$ is contained in an interior subset of the domain of attraction of $\bar{y}$.*

Then we have

$$\begin{array}{l} \lim_{\epsilon\to 0} x_\epsilon(t) = \bar{x}(t),\ 0 \le t \le L \\ \lim_{\epsilon\to 0} y_\epsilon(t) = \bar{y}(t),\ 0 \le d \le t \le L \end{array}$$

with d and L constants independent of ϵ.

In order to apply Tikhonov's theorem directly to the system Σ, we first need to rescale time according to $t \to \epsilon t$. This gives

$$\begin{array}{rcl} \epsilon\dfrac{dV}{dt} & = & -V + W \cdot S(V) + I \\ \dfrac{dW}{dt} & = & S(V) \otimes S(V) - \mu W. \end{array}$$

Tikhonov's theorem then implies that solutions of Σ are close to solutions of the reduced system (in the unscaled time variable)

$$\left\{ \begin{array}{rcl} V(t) & = & W \cdot S\big(V(t)\big) + I(t) \\ \dot{W} & = & \epsilon\Big(S(V) \otimes S(V) - \mu W\Big), \end{array} \right. \tag{10}$$

provided that the dynamical systems Σ in equation (6), and equation (10) are well defined. It is easy to show that both systems are Lipschitz because of the properties of S. Following [Faugeras et al., 2008], we know that if

$$S'_m\|W\| < 1, \tag{11}$$

then there exists an isolated root $\bar{V} : \mathbb{R}_+ \to \mathbb{R}^N$ of the equation $V = W \cdot S(V) + I$ and $\bar{V}$ is asymptotically stable. Equation (11) corresponds to the weakly connected case. Moreover, the initial condition belongs to the basin of attraction of this single fixed point. Note that we require $\frac{M}{T} \ll 1$ so that the membrane potentials have sufficient time to approach the equilibrium associated with a given input before the next input is presented to the network. In fact, this assumption make it reasonable to neglect the transient activity dynamics due to the switching between inputs.

3.2 Periodic averaging

The system given by equation (10) corresponds to a differential equation for W with T-periodic forcing due to the presence of V on the right–hand side. Since $T \ll \epsilon^{-1}$, we can use classical averaging methods to show that solutions of (10) are close to solutions of the following autonomous system on the time-interval $[0, \frac{1}{\epsilon}]$ (which we suppose large because $\epsilon << 1$)

$$\Sigma_0 : \quad \left\{ \begin{array}{lcl} V(t) & = & W \cdot S(V(t)) + I(t) \\ \\ \dfrac{dW}{dt} & = & \epsilon\Big(\dfrac{1}{T}\displaystyle\int_0^T S(V(s)) \otimes S(V(s))ds - \mu W(t)\Big). \end{array} \right. \tag{12}$$

It follows that solutions of Σ are also close to solutions of Σ_0. Finding the explicit solution $V(t)$ for each input $I(t)$ is difficult and requires fixed points methods, e.g. a Picard algorithm. Therefore, we will consider yet another system Σ' whose solutions are also close to Σ_0 and hence Σ. In order to construct Σ' we need to introduce some additional notation.

Let us label the M inputs by $I^{(a)}, a = 1, \ldots, M$ and denote by $V^{(a)}$ the fixed point solution of the equation $V^{(a)} = W \cdot S(V^{(a)}) + I^{(a)}$. Given the periodic sampling of the inputs, we can rewrite (12) as

$$\begin{aligned} V^{(a)} &= W \cdot S(V^{(a)}) + I^{(a)} \\ \frac{dW}{dt} &= \epsilon\Big(\frac{1}{M}\sum_{a=1}^{M} S(V^{(a)}) \otimes S(V^{(a)}) - \mu W(t)\Big). \end{aligned} \tag{13}$$

If we now introduce the $N \times M$ matrices $\mathcal{V}$ and $\mathcal{I}$ with components $\mathcal{V}_{ia} = V_i^{(a)}$ and $\mathcal{I}_{ia} = I_i^{(a)}$, then we can eliminate the tensor product and simply write (13) in the matrix form

$$\begin{aligned} \mathcal{V} &= W \cdot S(\mathcal{V}) + \mathcal{I} \\ \frac{dW}{dt} &= \epsilon\Big(\frac{1}{M} S(\mathcal{V}) \cdot S(\mathcal{V})^T - \mu W(t)\Big), \end{aligned} \tag{14}$$

where $S(\mathcal{V}) \in \mathbb{R}^{N\times M}$ such that $[S(\mathcal{V})]_{ia} = s(V_i^{(a)})$. A second application of Tikhonov's theorem (in the reverse direction) then establishes that solutions of the system Σ_0 (written in the matrix form (14)) are close to solutions of the matrix system

$$\Sigma' : \quad \begin{cases} \dfrac{d\mathcal{V}}{dt} &= -\mathcal{V} + W \cdot S\big(\mathcal{V}\big) + \mathcal{I} \\ \dfrac{dW}{dt} &= \epsilon\Big(\dfrac{1}{M} S(\mathcal{V}) \cdot S(\mathcal{V})^T - \mu W(t)\Big) \end{cases} \tag{15}$$

In the remainder of the paper we will focus on the system Σ' whose solutions are close to those of the original system Σ provided condition (11) is satisfied, i.e. the network is weakly connected. Clearly, the fixed points (V^*, W^*) of system Σ satisfy $\|W^*\| \leq \frac{S_m^2}{\mu}$. Therefore, equation (11) says that if $\frac{S_m^2 S_m'}{\mu} < 1$ then Tikhonov's theorem can be applied and systems Σ and Σ' can be reasonably considered as good approximations of each other. The advantage of the averaged system Σ' is that is given by autonomous ordinary differential equations. Moreover, since it is Lipschitz continuous, it leads to a well-posed Cauchy problem. Finally, note that it is straighforward to extend our approach to time-functional rules (e.g. sliding threshold or BCM rules as described in [Bienenstock et al., 1982]) which, in this new framework, would be approximated by simple ordinary differential equations (as opposed to time-functional differential equations) provided S is redefined appropriately.

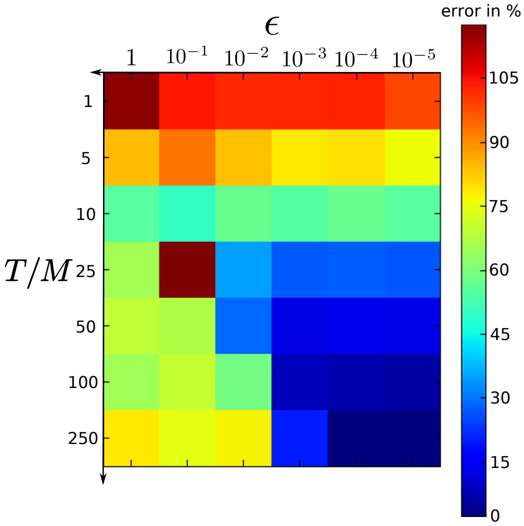

Figure 1: Percentage of error between final connectivities for the exact and averaged system.

3.3 Simulations

To illustrate the above approximation, we simulate a simple network with both exact, i.e. Σ, and averaged ,i.e. Σ', evolution equations. For these simulations, the network consists of $N = 10$ fully-connected neurons and is presented with $M = 10$ different random inputs taken uniformly in the intervals $[0, 1]^N$. For this simulation we use $s(x) = \frac{1}{1+e^{-4(x-1)}}$, and $\mu = 10$. Figure 1. shows the percentage of error between final connectivities for different values of ϵ and T/M. Figure 2 shows the temporal evolution of the norm of the connectivity for both the exact and averaged system for $T = 10^3$ and $\epsilon = 10^{-3}$.

4 Stability

4.1 Liapunov function

In the case of a single fixed input ($M = 1$), the systems Σ and Σ' are equivalent and reduce to the neural network with adapting synapses previously analyzed by [Dong and Hopfield, 1992]. Under the additional constraint that the weights are symmetric ($W_{ij} = W_{ji}$), these authors showed that the simultaneous evolution of the neuronal activity variables and the synaptic weights can be re-expressed as a

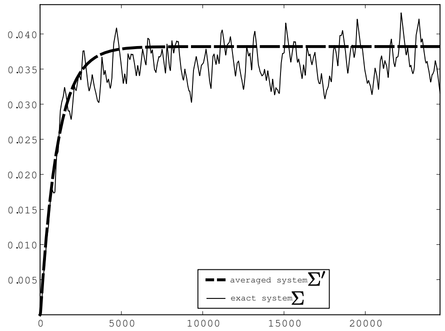

Figure 2: Temporal evolution of the norm of the connectivities of the exact system Σ and averaged system Σ'.

gradient dynamical system that minimizes a Liapunov or energy function of state. We can generalize their analysis to the case of multiple inputs ($M > 1$) and non-symmetric weights using the averaged system Σ'. That is, following along similar lines to [Dong and Hopfield, 1992], we introduce the energy function

$$E(\mathcal{U}, W) = -\frac{1}{2}\langle \mathcal{U}, W \cdot \mathcal{U} \rangle - \langle \mathcal{I}, \mathcal{U} \rangle + \langle 1, \overline{S^{-1}}(\mathcal{U}) \rangle + \frac{M\mu}{2}\|W\|^2 \qquad (16)$$

where $\mathcal{U} = S(\mathcal{V})$, $\|W\|^2 = \langle W, W \rangle = \sum_{i,j} W_{ij}^2$,

$$\langle \mathcal{U}, W \cdot \mathcal{U} \rangle = \sum_{a=1}^{M}\sum_{i=1}^{N} U_i^{(a)} W_{ij} U_j^{(a)}, \quad \langle \mathcal{I}, \mathcal{U} \rangle = \sum_{a=1}^{M}\sum_{i=1}^{N} I_i^{(a)} U_i^{(a)} \qquad (17)$$

and

$$\langle 1, \overline{S^{-1}}(\mathcal{U}) \rangle = \sum_{a=1}^{M}\sum_{i=1}^{N} \int_0^{U_i^{(a)}} S^{-1}(\xi) d\xi. \qquad (18)$$

In contrast to [Dong and Hopfield, 1992], we do not require *a priori* that the weight matrix is symmetric. However, it can be shown that the system always converges to a symmetric connectivity pattern. More precisely, $\mathcal{A} = \Big\{(\mathcal{V}, W) \in \mathbb{R}^{N\times M} \times \mathbb{R}^{N\times N} : W = W^T\Big\}$ is an attractor of the system Σ'. A proof can be found in appendix 8.1. It can then be shown that on $\mathcal{A}$ (symmetric weights), E is a Liapunov function of the dynamical system Σ', that is,

$$\frac{dE}{dt} \leq 0, \quad \text{and} \quad \frac{dE}{dt} = 0 \implies \frac{d\mathcal{Y}}{dt} = 0, \quad \mathcal{Y} = (\mathcal{V}, W)^T.$$

The boundedness of E and the Krasovskii-LaSalle invariance principle then implies that the system converges to an equilibrium [Khalil and Grizzle, 1996]. We thus have

Theorem 4.1. *The initial value problem for the system Σ' with $\mathcal{Y}(0) \in \mathcal{H}$, converges to an equilibrium state.*

Proof. See appendix 8.2 □

It follows that neither oscillatory nor chaotic attractor dynamics can occur.

4.2 Linear stability

Although we have shown that there are stable fixed points, not all of the fixed points are stable. However, we can apply a linear stability analysis on the system Σ' to derive a simple sufficient condition for a fixed point to be stable. The method we use in the proof could be extended to more complex rules. The proof reveals the significant role played by the Kronecker product in Hebbian learning.

Theorem 4.2. *The equilibria of system Σ' satisfy:*

$$\left\{ \begin{array}{l} \mathcal{V}^* = \frac{1}{\mu M} S(\mathcal{V}^*) \cdot S(\mathcal{V}^*)^T \cdot S(\mathcal{V}^*) + \mathcal{I} \\ W^* = \frac{1}{\mu M} S(\mathcal{V}^*) \cdot S(\mathcal{V}^*)^T \end{array} \right. \tag{19}$$

and a sufficient condition for stability is

$$3S'_m \|W^*\| < 1 \tag{20}$$

provided $1 > \epsilon\mu$ *which is most probably the case since* $\epsilon << 1$.

Proof. See appendix 8.3 □

This condition is strikingly similar to that derived in [Faugeras et al., 2008]. In fact, condition (20) is stronger than the contracting condition (11). It says the network may converge to a weakly connected situation. It justifies the averaging method by saying that we remain in the domain of validity of the averaging method. It also says that the dynamics of $\mathcal{V}$ is likely (because the condition is only sufficient) to be contracting and therefore subject to no bifurcations: a fully recurrent learning neural network is likely to have a “simple” dynamics.

5 Geometrical structure of equilibrium points

5.1 Learning the correlation matrix of the inputs

It follows from equation (19) that the equilibrium weight matrix W^* is given by the correlation matrix of the firing rates. Moreover, in the case of sufficiently large inputs, the matrix of equilibrium membrane potentials satisfies $\mathcal{V}^* \approx \mathcal{I}$. More precisely, if $|S(I_i^{(a)})| \ll |I_i^{(a)}|$ for all $a = 1, \ldots, M$ and $i = 1, \ldots, N$, then we can generate an iterative solution for $\mathcal{V}^*$ of the form

$$\mathcal{V}^* = \mathcal{I} + \frac{1}{\mu} S\big(\mathcal{I}\big) \cdot S\big(\mathcal{I}\big)^T \cdot S\big(\mathcal{I}\big) + \mathrm{h.o.t.}$$

On the other hand, if the inputs are comparable in size to the synaptic weights, then there is no explicit solution for $\mathcal{V}^*$. Roughly speaking, we observe that the connection term has the role of "smoothing" the solution. Therefore, if a Gaussian is presented to the network (as the only input), the membrane potential is likely to be another Gaussian with a larger variance. If no input is presented to the network ($I = 0$), then $S(0) \neq 0$ implies that the activity is non-zero, that is, there is spontaneous activity. Combining these observations, we see that the network roughly extracts and stores the correlation matrix of the strongest inputs within the weights of the network.

5.2 From a symmetric connectivity matrix to a convolutional network

So far neurons have been identified by a label $i \in \{1..N\}$; there is no notion of geometry or space in the preceding results. However, as we show below, the inputs may contain a spatial structure that can be encoded by the connectivity. In this section, we propose a mechanism to unveil the hidden geometrical structure within the connectivity. More specifically, we want to find an integer $k \in \mathbb{N}$ and N points in $\mathbb{R}^k$, denoted by $x_i, i \in \{1, \ldots, N\}$, so that the connectivity can roughly be written as $W_{ij}^* \simeq \exp(-\|x_i - x_j\|^2)$. In other words, we interpret the final connectivity as a matrix describing the distance between the neurons living in a k-dimensional space. However, W^* is not always a distance matrix, therefore, it is natural to project it on the set of distance matrices. Finding the best fit of W^* to a distance matrix is usually called multidimensional scaling. This set of methods is reviewed in [Borg and Groenen, 2005].

First, define $\widehat{W} \in \mathbb{R}^{N\times N}$ as W^* whose diagonal terms are set to W^*_{max} the largest component of W^*: $W^*_{ij} = N_{ij}\widehat{W}_{ij}$ with $N_{ij} = 1$ if $i \neq j$ and $N_{ii} = W_{ii}/W^*_{max}$. Second, define a bijective kernel function on $x \in \mathbb{R}_+$ such that $K(x) = W^*_{max}e^{-x/\sigma^2}$. Given that $\widehat{W}$ is non-negative, we define the matrix $D = K^{-1}(\widehat{W})$ corresponding to the application of the inverse of K to each component of $\widehat{W}$. As said before, we want to find $k \in \mathbb{N}$ and $x_i \in \mathbb{R}^k$ for $i \in \{1,\ldots,N\}$ so that D is a distance matrix, $D_{ij} = \|x_i - x_j\|^2$. In general, this is not possible. However, we can compute the projection of the symmetric matrix $\widehat{W}$ onto the set of distance matrices by applying multidimensional scaling methods as described in [Borg and Groenen, 2005]. We use the stress majorization or SMACOF algorithm for the stress1 cost function throughout the article. In other words, we can find the distance matrix $D_\perp$ such that $\|D_\shortparallel\| = \|D - D_\perp\|$ is minimal. Therefore, $\widehat{W}_{ij} = K(D_\shortparallel)_{ij}K(D_\perp)_{ij}$. Define M such that $M(x_i,x_j) = K(D_\shortparallel)_{ij}\, N_{ij}$ and G_σ a Gaussian with a standard deviation equal to σ. In spatial coordinates

$$W^*(x_i,x_j) = M(x_i,x_j)\; G_\sigma(\|x_i - x_j\|) \tag{21}$$

Multidimensional scaling methods consist in minimizing the contribution of M in the preceding equation. Hence, we refer to M as the non-convolutional connectivity.

A position $x_i \in \mathbb{R}^k$ is associated to each neuron $i \in \{1,\ldots,N\}$ such that

$$\big(W \cdot S(V)\big)_i = \sum_{j=1}^{N} W^*_{ij}S(V_j) = \sum_{j=1}^{N} M(x_i,x_j)\; G_\sigma(\|x_i - x_j\|)\; S(V(x_j)) \tag{22}$$

In particular, we can assume that k is large enough for $\|D_\shortparallel\|$ to be very small. Moreover, if the neurons are equally excited on average (i.e. the diagonal of W^* is already equal to $W^*_{max}I_d$), then it is reasonable to consider that $M(x_i,x_j) = 1$ leading to the following convolutional product

$$W \cdot S(V) = G_\sigma(\|.\|) * S(V)$$

Therefore, in the space defined by the $x_i \in \mathbb{R}^k$ the connectivity is close to being convolutional.

5.3 Unveiling the geometrical structure of the inputs

We hypothesize that the space defined by the x_i reflects the underlying geometrical structure of the inputs. We have not found a way to prove this so we only provide numerical examples that illustrate this claim. In the following examples, we relate the geometry of the manifold suggested by the x_i to the network inputs. Thus we feed the network with inputs having a defined geometrical structure and then show how this structure can be extracted from the connectivity by the method above. However, it is by extracting the structure from unknown inputs that these networks might reveal themselves useful. Therefore, the following is only a (numerical) proof of concept.

We assume the inputs to be uniformly distributed over a manifold Ω with fixed geometry. This strong assumption amounts to considering that the feedforward connectivity (which we do not consider here) has already properly filtered the information coming from the sensory organs. More precisely, define the set of inputs by the matrix $\mathcal{I} \in \mathbb{R}^{N\times M}$ such that $I_i^{(a)} = f(\|y_i - z_a\|_\Omega)$ where the z_a are uniformly distributed points over Ω, the y_i are the positions on Ω that "label" the ith neuron, and f is a decreasing function on $\mathbb{R}_+$. The norm $\|.\|_\Omega$ is the natural norm defined over the manifold Ω. For simplicity, assume $f(x) = f_\sigma(x) = Ae^{-\frac{x^2}{\sigma^2}}$ so that the inputs are localized bell-shaped bumps on the shape Ω.

5.3.1 Planar retinotopy

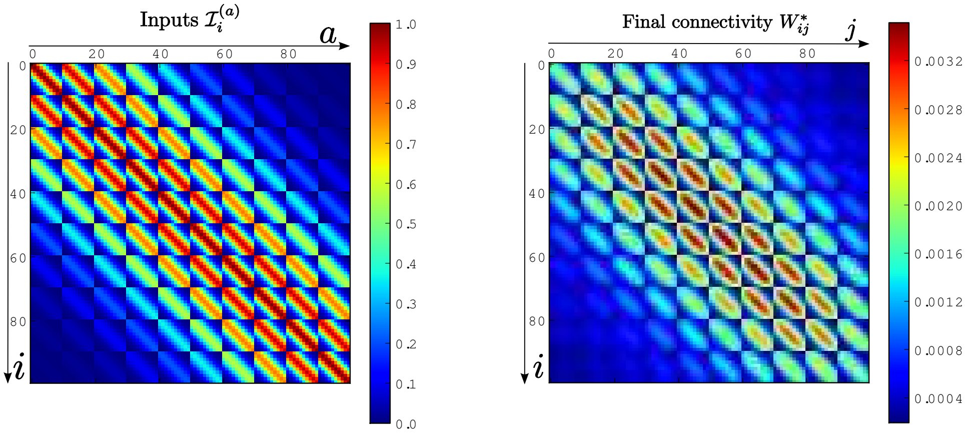

Figure 3: Plot of planar retinotopic inputs on $\Omega = [0,1]\times[0,1]$ (left) and final connectivity matrix of the system Σ' (right). The parameters used for this simulation are $s(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\mu = 10$, $\epsilon = 0.001$, $N = M = 100$, $\sigma = 4$.

We consider a set of Gaussian inputs uniformly distributed over a two-dimensional plane, e.g. $\Omega = [0,1] \times [0,1]$. For simplicity, we take $N = M = K^2$ and set $z_a = y_i$ for $i = a$, $a \in \{1, \dots, M\}$. (The numerical results show an identical structure for the final connectivity when the y_j correspond to random points, but the analysis is harder). In the simpler case of one-dimensional Gaussians with $N = M = K$, the input matrix takes the form $\mathcal{I} = T_{f_\sigma}$, where T_f is a symmetric Toeplitz matrix:

$$T_f = \begin{pmatrix} f(0) & f(1) & f(2) & \cdots & \cdots & f(K) \\ f(1) & f(0) & f(1) & f(2) & \cdots & f(K-1) \\ f(2) & f(1) & f(0) & f(1) & \cdots & f(K-2) \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ f(K) & f(K-1) & f(K-2) & \cdots & \cdots & f(0) \end{pmatrix} \tag{23}$$

In the two-dimensional case, we set $y = (u, v) \in \Omega$ and introduce the labeling $y_{k+(l-1)K} = (u_k, v_l)$ for $k, l = 1, \dots K$. It follows that $I_i^{(a)} \sim \exp(-(u_k - u_{k'})^2) \exp(-(v_l - v_{l'})^2)$ for $i = k + (l-1)K$ and $a = k' + (l'-1)K$. Hence, we can write $\mathcal{I} = T_{f_\sigma} \otimes T_{f_\sigma}$, where $\otimes$ is the Kronecker product; the Kronecker product is responsible for the $K \times K$ sub-structure we can observe in figure 3 with $K = 10$. Note that if we were interested in a n-dimensional retinotopy, then the input matrix could be written as a Kronecker product between n Toeplitz matrices. As previously mentioned, the final connectivity matrix roughly corresponds to the correlation matrix of the input matrix. It turns out that the correlation matrix of $\mathcal{I}$ is also a Kronecker product of two Toeplitz matrix generated by a single Gaussian (with a different standard deviation). Thus, the connectivity matrix has the same basic form as the input matrix when $z_a = y_i$ for $i = a$. The inputs and stable equilibrium points of the simulated system are shown in figure 3. The positions x_i of the neurons after multidimensional scaling are shown in figure 4.

5.3.2 Toroïdal retinotopy

We now assume that the inputs are uniformly distributed over a two-dimensional torus, i.e. $\Omega = \mathbb{T}^2$. That is, the input labels z_a are randomly distributed on the torus. The neuron labels y_i are regularly and uniformly distributed on the torus. The inputs and final stable weight matrix of the simulated system are shown in figure 5. The positions x_i of the neurons after multidimensional scaling for $k = 3$ are shown in figure 6, and appear to form a cloud of points distributed on a torus. In order to confirm this, we have used a numerical method from computational cohomology [Zomorodian and Carlsson, 2005] to construct the so–called persistent

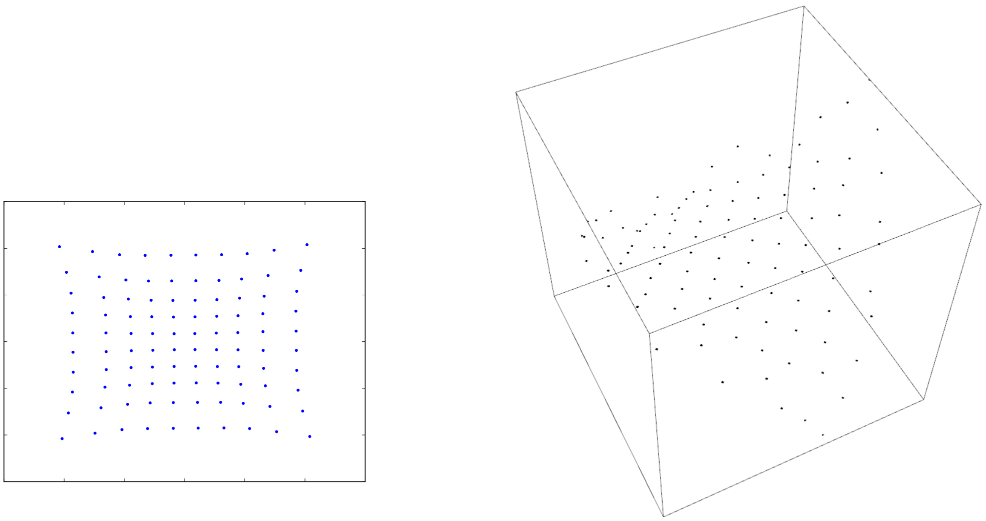

Figure 4: Positions x_i of the neurons after having applied classical multidimensional scaling to the final connectivity matrix shown in figure 3 for $k = 2$ (left) and $k = 3$ (right). The regular spacing of the neurons for $k = 2$ shows that the planar structure of the inputs has been recovered, although the corner of the square appear less regular due to boundary effects. In the case $k = 3$, there is an embedding of the plane into three dimensions; the saddle–like shape accounts for the corner irregularity observed when $k = 2$.

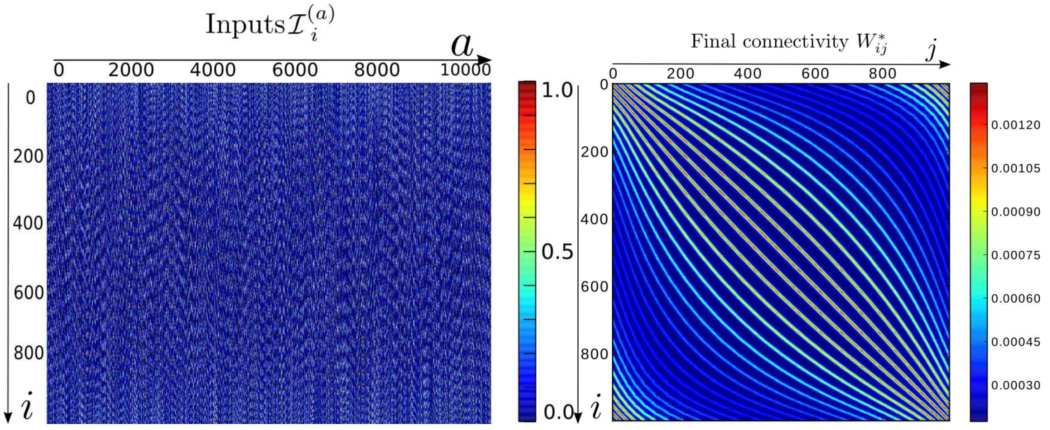

Figure 5: Plot of retinotopic inputs on $\Omega = \mathbb{T}^2$ (left) and the final connectivity matrix (right) for the system Σ'. The parameters used for this simulation are $s(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\mu = 10$, $\epsilon = 0.001$, $N = 1000, M = 10,000, \sigma = 2$.

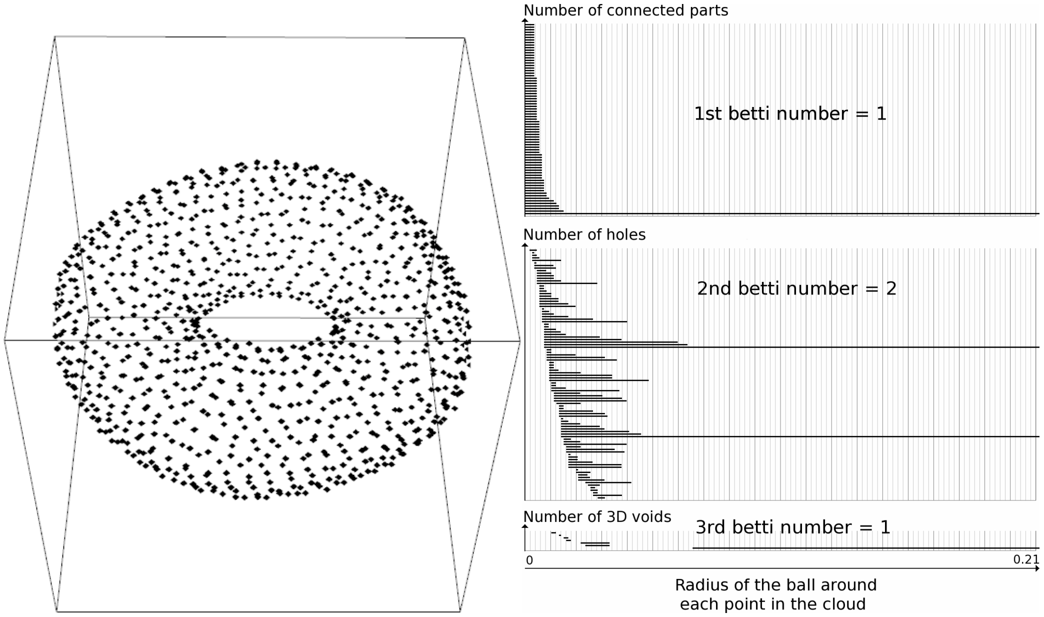

Figure 6: Left: Positions x_i of the neurons for $k = 3$ after having applied multidimensional scaling methods presented in part 5.2 to the final connectivity matrix shown in figure 5. Right: Persistent cohomology barcodes of the cloud of points x_i computed using the Jplex software package of [Sexton and Vejdemo-Johansson,]. (See section 8.4 for a short introduction to persistent cohomology). The triplet of betti numbers (1,2,1) appear stable confirming that the points lie on a 2 dimensional torus.

cohomology barcodes of the neurons' positions. These determine certain topological invariants of the underlying space. (See section 8.4 for a short introduction to persistent cohomology and barcodes). The results are also shown in figure 6, and establish that the network has learnt the underlying toroidal geometry of the inputs.

5.4 Links with neuroanatomy

The brain is subject to energy constraints which are completely neglected in the above formulation. These constraints most likely have a significant impact on the positions of real neurons in the brain. Indeed, it seems reasonable to assume that the positions and connections of neurons reflect a trade-off between the energy costs of biological tissue and their need to process information effectively. For instance, it has been suggested that a principle of wire length minimization may occur in the brain [Swindale, 1996, Chklovskii et al., 2002]. In our neural mass framework, one may consider that the stronger two neural masses are connected, the larger the number of real axons linking the neurons together. Therefore, minimizing axonal length can be read as: the stronger the connection the closer, which is consistent with the convolutional part of the weight matrix. However, the underlying geometry of natural inputs is likely to be very high-dimensional, whereas the brain lies in a three-dimensional world. In fact, the cortex is so flat that it is effectively two-dimensional. Hence, the positions of real neurons are different from the positions $x_i \in \mathbb{R}^k$ in a high dimensional vector space; since the cortex is roughly two-dimensional, the positions could only be realized physically if $k = 2$. Therefore, the three-dimensional toric geometry or any higher dimensional structure could not be perfectly implemented in the cortex without the help of non-convolutional long-range connections. Indeed, we suggest that the cortical connectivity is made of two parts: i) a local convolutional connectivity corresponding to the convolutional term G_σ in (21), which is consistent with the requirements of energy efficiency, and ii) a non-convolutional connectivity corresponding to the factor M in equation (21), which is required in order to represent various stimulus features. If the cortex were higher-dimensional ($k \gg 2$) then $M \equiv 1$.

We illustrate the above claim by considering two examples based on the functional anatomy of the primary visual cortex: the emergence of ocular dominance columns and orientation columns, respectively. We proceed by returning to the case of planar retinotopy (section 5.3.1) but now with additional input structure. In the first case, the inputs are taken to be binocular and isotropic, whereas in the second case they are taken to be monocular and anisotropic. The details are presented below. Given a set of prescribed inputs, the network evolves according to equation (15) and the lateral connections converge to a stable equilibrium. The resulting weight matrix is then projected onto the set of distance matrices for $k = 2$ (as described in section 5.2) using the stress majorization or SMACOF algorithm for the stress1 cost function as described in [Borg and Groenen, 2005]. We thus assign a position $x_i \in \mathbb{R}^2$ to the ith neuron, $i = 1, \ldots, N$. (Note that the position x_i extracted from the weights using multidimensional scaling is distinct from the "physical" position y_i of the neuron in the retinocortical plane; the latter determines the center of its receptive field). The convolutional connectivity (G_σ in equation 21) is therefore completely defined: on the planar map of points x_i, neurons are isotropically connected to their neighbors; the closer the neurons are the stronger is their convolutional connection. Moreover, since the stimulus feature preferences (orientation, ocular dominance) of each neuron i, $i = 1, \ldots, N$, are prescribed, we can superimpose these feature preferences on to the planar map of points x_i. In both examples, we find that neurons with the same ocular or orientation selectivity tend to cluster together (see figures 7 and 8): interpolating these clusters then generates corresponding feature columns. It is important to emphasize that the retinocortical positions y_i do not have any columnar structure, that is, they do not form clusters with similar feature preferences. Thus, in contrast to standard developmental models of vertical connections, the columnar structure emerges from the recurrent weights following Hebbian learning and an application of multidimensional scaling. It follows that neurons coding for the same feature tend to be strongly connected; indeed, the multidimensional scaling algorithm has the property that it positions strongly connected neurons close together . Equation (21) also suggests that the connectivity has a non-convolutional part, M, which is a consequence of the low-dimensionality ($k = 2$). In order to illustrate the structure of the non-convolutional connectivity, we select a neuron i in the plane and draw a link from it at position x_i to the neurons at position x_j for which $M(x_i, x_j)$ is maximal. We find that M tends to be patchy, i.e. it connects neurons having the same feature preferences. In the case of orientation, M also tends to be co-aligned, i.e. connecting neurons with similar orientation preference along a vector in the plane of the same orientation.

5.4.1 Ocular dominance columns and patchy connectivity

In order to construct binocular inputs, we partition the N neurons into two sets $i \in \{1, \ldots, N/2\}$ and $i \in \{N/2 + 1, \ldots, N\}$ that code for the left and right eyes, respectively. The ith neuron is then given a retinocortical position $y_i \in [0, 1] \times$

$[0,1]$, with the y_i uniformly distributed across the plane. We do not assume *a priori* that there exist any ocular dominance columns, that is, neurons with similar retinocortical positions y_i do not form clusters of cells coding for the same eye. We then take the ath input to the network to be of the form

$$\begin{aligned}
I_i^{(a)} &= (1+\gamma(a))e^{-\frac{(y_i-z_a)^2}{\sigma'^2}}, \quad i = 1, \ldots, N/2 \\
I_i^{(a)} &= (1-\gamma(a))e^{-\frac{(y_i-z_a+s)^2}{\sigma'^2}}, \quad i = N/2+1, \ldots, N,
\end{aligned}$$

where $s \in \mathbb{R}^2$ represents some form of binocular disparity, z_a and $\gamma(a)$ are randomly generated from $[0,1]^2$ and $[-1,1]$, respectively, see [Bressloff, 2005]. Thus, if $\gamma(a) > 0$ ($\gamma(a) < 0$) then the corresponding input is predominantly from the left (right) eye. In our simulations we take $\sigma = \sigma' = 0.1$ and $s = 0.005$. The results of our simulations are shown in figure 7. In particular, we plot the points x_i obtained by performing multidimensional scaling on the final connectivity matrix for $k = 2$, and superimposing upon this the ocular dominance map obtained by interpolating between clusters of neurons with the same eye preference. We also illustrate the non-convolutional connectivity by linking one selected neuron to the five neurons labeled j it is most strongly connected to (with $M(x_i, x_j) > 1$), with i the label of the central neuron. This clearly shows that long–range connections tend to link cells with the same ocular dominance.

5.4.2 Orientation columns and colinear connectivity

In order to construct oriented inputs, we partition the N neurons into four groups Σ_θ corresponding to different orientation preferences $\theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. Thus, if neuron $i \in \Sigma_\theta$ then its orientation preference is $\theta_i = \theta$. For each group, the neurons are randomly assigned a retinocortical position $y_i \in [0,1] \times [0,1]$. Again, we do not assume *a priori* that there exist any orientation columns, that is, neurons with similar retinocortical positions y_i do not form clusters of cells coding for the same orientation preference. Each cortical input $I_i^{(a)}$ is generated by convolving a thalamic input consisting of an oriented Gaussian with a Gabor–like receptive field [Miikkulainen et al., 2005]. Let $\mathcal{R}_\theta$ denote a 2-dimensional rigid body rotation in the plane with $\theta \in [0, 2\pi)$. Then

$$I_i^{(a)} = \int G_i(\xi - y_i) I_a(\xi - z_a) d\xi, \tag{24}$$

where

$$G_i(\xi) = G_0(\mathcal{R}_{\theta_i}\xi) \tag{25}$$

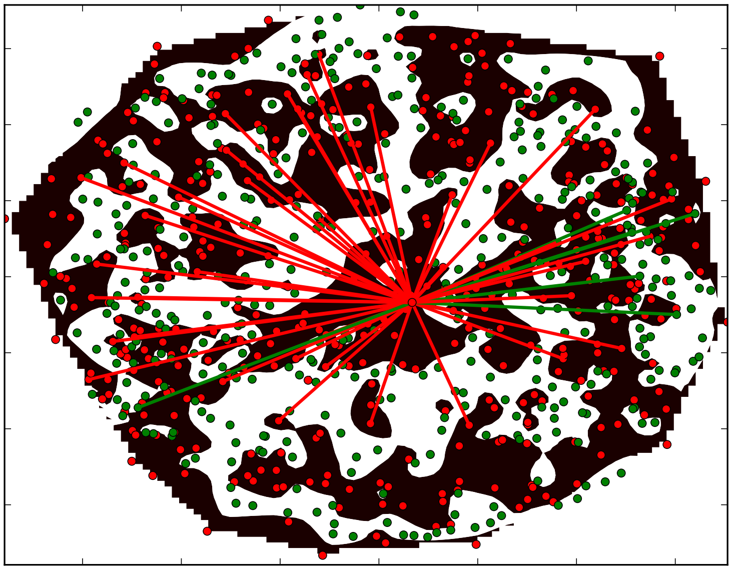

Figure 7: Plot of the positions x_i of neurons for $k = 2$ in red (right eye) and green (left eye). We have used an interpolation method to highlight the areas dominated by the right eye in black. These ocular dominance columns have fractal borders which are less regular than those observed in optical imaging experiments. The convolutional connectivity (G_σ in equation (21)) is implicitly described by the position of the neurons: the closer the neurons, the stronger their connections. The strongest components of the non-convolutional connectivity (M in equation (21)) from a central red neuron are also shown by drawing links from this neuron to the target neurons. The color of the link refers to the color of the target neuron. Therefore, we see that it is mainly connected to neurons of its same ocular dominance resulting in a patchy distribution. The parameters used for this simulation are $s(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\mu = 10$, $\epsilon = 0.01$, $N = 800$ $M = 3200$.

and $G_0(\xi)$ is the Gabor–like function

$$G_0(\xi) = A_+ e^{-\xi^T.\Lambda^{-1}.\xi} - A_- e^{-(\xi - e_0)^T.\Lambda^{-1}.(\xi - e_0)} - A_- e^{-(\xi + e_0)^T.\Lambda^{-1}.(\xi + e_0)}$$

with $e_0 = (0, 1)$ and

$$\Lambda = \begin{pmatrix} \sigma_{large} & 0 \\ 0 & \sigma_{small} \end{pmatrix}.$$

The amplitudes A_+, A_- are chosen so that $\int G_0(\xi) d\xi = 0$. Similarly, the thalamic input $I_a(\xi) = I(\mathcal{R}_{\theta'_a}\xi)$ with $I(\xi)$ the anisotropic Gaussian

$$I(\xi) = e^{-\xi^T.\Lambda'^{-1}.\xi}, \qquad \Lambda' = \begin{pmatrix} \sigma'_{large} & 0 \\ 0 & \sigma'_{small} \end{pmatrix}.$$

The input parameters θ'_a and z_a are randomly generated from $[0, \pi)$ and $[0, 1]^2$ respectively. In our simulations we take $\sigma_{large} = 0.133...$, $\sigma'_{large} = 0.266...$ and $\sigma_{small} = \sigma'_{small} = 0.0333...$. The results of our simulations are shown in figure 8. In particular, we plot the points x_i obtained by performing multidimensional scaling on the final connectivity matrix for $k = 2$, and superimposing upon this the orientation preference map obtained by interpolating between clusters of neurons with the same orientation preference. To avoid border problems we have zoomed on the center on the map. We also illustrate the non-convolutional connectivity by linking one selected neuron to all other neurons for which M is maximal. The patchy, anisotropic nature of the long–range connections is clearly seen. The anisotropic nature of the connections is further quantified in the histogram of figure 9.

6 Discussion

In this paper, we have shown how a neural network can learn the underlying geometry of a set of inputs. We have considered a fully recurrent neural network whose dynamics is described by a simple non-linear rate equation, together with unsupervised Hebbian learning with decay that occurs on a much slower time scale. Although several inputs are periodically presented to the network, so that the resulting dynamical system is non-autonomous, we have shown that such a system has a fairly simple dynamics: the network connectivity matrix always converges to an equilibrium point. We have then demonstrated how this connectivity matrix can be expressed as a distance matrix in $\mathbb{R}^k$ for sufficiently large k, which can be

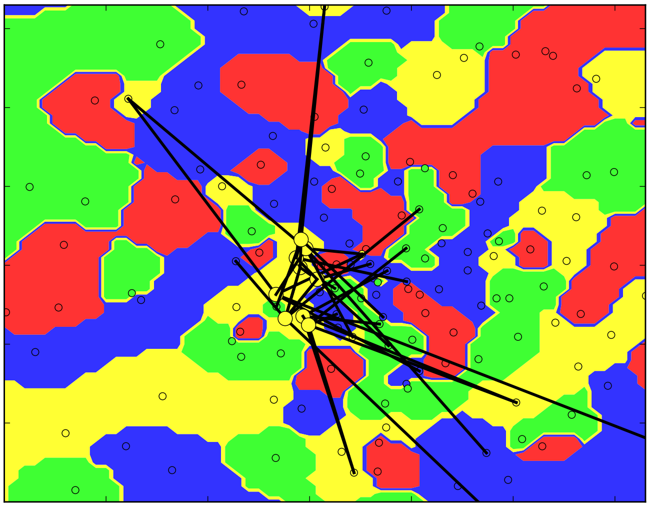

Figure 8: Plot of the positions x_i of neurons for $k = 2$ obtained by multidimensional scaling of the weight matrix. Neurons are clustered in orientation columns represented by the colored areas, which are computed by interpolation. The strongest components of the non-convolutional connectivity (M in equation (21)) from a particular neuron in a yellow area are illustrated by drawing black links from this neuron to the target neurons. Since the yellow color corresponds to an orientation of $\frac{3\pi}{4}$, the non-convolutional connectivity shows the existence of a co-linear connectivity as exposed in [Bosking et al., 1997]. The parameters used for this simulation are $s(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\mu = 10$, $\epsilon = 0.01$, $N = 900$ $M = 9000$.

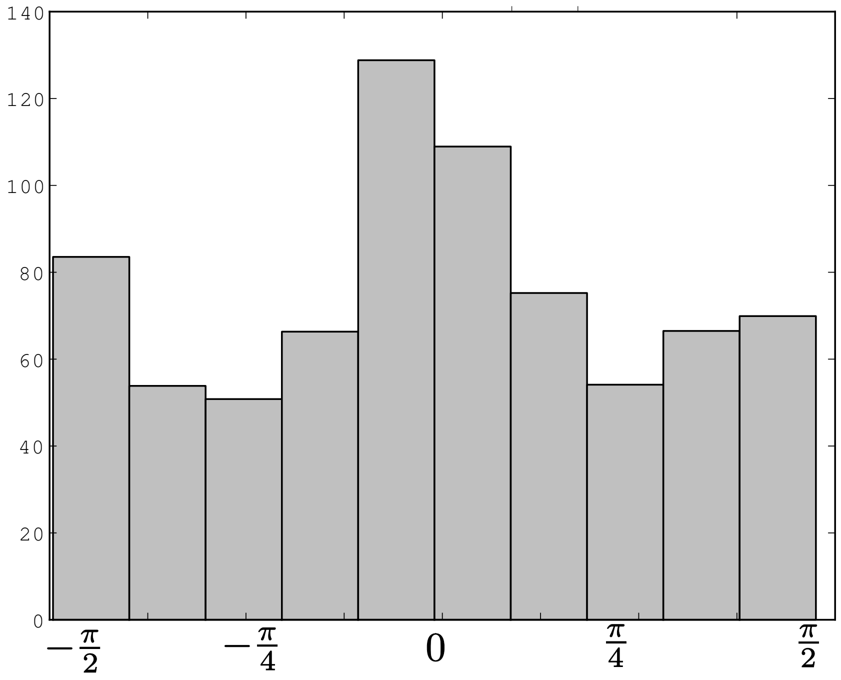

Figure 9: Histogram of the 5 largest components of the non-convolutional connectivity for 80 neurons randomly chosen among those shown in Fig. 8. The abcissa corresponds to the difference in radian between the direction preference of the neuron and the direction of the links between the neuron and the target neurons. This histogram is weighted by the strength of the non-convolutional connectivity. It shows a preference for co-aligned neurons but also a slight preference for perpendicularly-aligned neurons (e.g. neurons of the same orientation but parallel to each other).

related to the underlying geometrical structure of the inputs. If the connectivity matrix is embedded in a lower two-dimensional space ($k = 2$), then the emerging patterns are similar to experimentally observed cortical feature maps. That is, neurons with the same feature preferences tend to cluster together forming cortical columns within the embedding space. Moreover, the recurrent weights decompose into a local isotropic convolutional part, which is consistent with the requirements of energy efficiency, and a longer–range non-convolutional part that is patchy. This suggest a new interpretation of the cortical maps: they correspond to two-dimensional embeddings of the underlying geometry of the inputs.

One of the limitations of applying simple Hebbian learning to recurrent cortical connections is that it only takes into account excitatory connections, whereas 20% of cortical neurons are inhibitory. Indeed, in most developmental models of feedforward connections, it is assumed that the local and convolutional connections in cortex have a Mexican hat shape with negative (inhibitory) lobes for neurons that are sufficiently far from each other. From a computational perspective, it is possible to obtain such a weight distribution by replacing Hebbian learning with some form of covariance learning ([Sejnowski and Tesauro, 1989]). However, it is difficult to prove convergence to a fixed point in the case of the covariance learning rule, and multidimensional scaling method cannot be applied directly unless the Mexican hat function is truncated so that it is invertible. Another limitation of rate-based Hebbian learning is that it does not take into account causality, in contrast to more biologically detailed mechanisms such as spike timing dependent plasticity.

The approach taken here is very different from standard treatments of cortical development [Miller et al., 1989, Swindale, 1996], in which the recurrent connections are assumed to be fixed and of convolutional Mexican hat form whilst the feedforward vertical connections undergo some form of correlation-based Hebbian learning. In the latter case, cortical feature maps form in the physical space of retinocortoical coordinates y_i, rather than in the more abstract planar space of points x_i obtained by applying multidimensional scaling to recurrent weights undergoing Hebbian learning in the presence of fixed vertical connections. A particular feature of cortical maps formed by modifiable feedforward connections is that the mean size of a column is determined by a Turing-like pattern forming instability, and depends on the length scales of the Mexican hat weight function and the two-point input correlations [Miller et al., 1989, Swindale, 1996]. No such Turing mechanism exists in our approach so that the resulting cortical maps tend to be more fractal-like (many length scales) compared to real cortical maps. Nevertheless, we have established that the geometrical structure of cortical feature maps can also be encoded by modifiable recurrent connections. This should have interesting consequences for models that consider the joint development of feed-forward and recurrent cortical connections. One possibility is that the embedding space of points x_i arising from multidimensional scaling of the weights becomes identified with the physical space of retinocortical positions y_i. The emergence of local convolutional structures together with sparser long-range connections would then be consistent with energy efficiency constraints in physical space.

Our paper also draws a direct link between the recurrent connectivity of the network and the positions of neurons in some vector space such as $\mathbb{R}^2$. In other words, learning corresponds to moving neurons or nodes so that their final position will match the inputs' geometrical structure. Similarly, the Kohonen algorithm [Kohonen, 1990] describes a way to move nodes according to the inputs presented to the network. It also converges toward the underlying geometry of the set of inputs. Although not formally equivalent, it seems that both of these approaches have the same qualitative behaviour. However, our method is more general in the sense that no neighborhood structure is assumed *a priori*; such a structure emerges via the embedding into $\mathbb{R}^k$.

Finally, note that we have used a discrete formalism based on a finite number of neurons. However, the resulting convolutional structure obtained by expressing the weight matrix as a distance matrix in $\mathbb{R}^k$, equation (21), allows us to take an appropriate continuum limit. This then generates a continuous neural field model in the form of an integro-differential equation whose integral kernel is given by the underlying weight distribution. Neural fields have been used increasingly to study large–scale cortical dynamics (see [Coombes, 2005] for a review). Our geometrical learning theory provides a developmental mechanism for the formation of these neural fields. One of the useful features of neural fields from a mathematical perspective is that many of the methods of partial differential equations can be carried over. Indeed, for a general class of connectivity functions defined over continuous neural fields, a reaction-diffusion equation can be derived whose solution approximates the firing rate of the associated neural field [Degond and Mas-Gallic, 1989, Cottet, 1995, Edwards, 1996]. It appears that the necessary connectivity functions are precisely those that can be written in the form (21). This suggests that a network that has been trained on a set inputs with an appropriate geometrical structure behaves as a diffusion equation in a high-dimensional space together with a reaction term corresponding to the inputs.

7 Acknowldegments

MNG and ODF were partially funded by the ERC advanced grant NerVi nb 227747. MG was partially funded by the région PACA, France. This publication was based on work supported in part by the National Science Foundation (DMS-0813677) and by Award No KUK-C1-013-4 made by King Abdullah University of Science and Technology (KAUST). PCB was also partially supported by the Royal Society Wolfson Foundation.

8 Appendix

8.1 Proof of the convergence to the symmetric attractor $\mathcal{A}$

We need to prove the 2 points: (i) $\mathcal{A}$ is an invariant set, and (ii) for all $\mathcal{Y}(0) \in \mathbb{R}^{N \times M} \times \mathbb{R}^{N \times N}$, $\mathcal{Y}(t)$ converges to $\mathcal{A}$ as $t \to +\infty$. Since $\mathbb{R}^{N \times N}$ is the direct sum of the set of symmetric connectivities and the set of anti-symmetric connectivies, we write $W(t) = W_S(t) + W_A(t),\ \forall t \in \mathbb{R}_+$, where W_S is symetric and W_A is anti-symetric.

(i) In (15), the right hand side of the equation for $\dot{W}$ is symmetric. Therefore, if $\exists t_1 \in R_+$ such that $W_A(t_1) = 0$, then W remains in $\mathcal{A}$ for $t \geq t_1$.

(ii) Projecting the expression for $\dot{W}$ in equation (15) on to the anti-symmetric component leads to

$$\frac{dW_A}{dt} = -\epsilon\mu W_A(t) \qquad (26)$$

whose solution is $W_A(t) = W_A(0)\exp(-\epsilon\mu t), \forall t \in \mathbb{R}_+$. Therefore, $\lim_{t \to +\infty} W_A(t) = 0$. The system converges exponentially to $\mathcal{A}$.

8.2 Proof of theorem 4.1

Consider the following Lyapunov function (see equation (16))

$$E(\mathcal{U}, W) = -\frac{1}{2}\langle \mathcal{U}, W \cdot \mathcal{U} \rangle - \langle \mathcal{I}, \mathcal{U} \rangle + \langle 1, \overline{S^{-1}}(\mathcal{U}) \rangle + \frac{\tilde{\mu}}{2}\|W\|^2, \qquad (27)$$

where $\tilde{\mu} = \mu M$, such that if $W = W_S + W_A$, where W_S is symmetric and W_A is anti-symmetric.

$$-\nabla E(\mathcal{U}, W) = \begin{pmatrix} W_S \cdot \mathcal{U} + I - S^{-1}(\mathcal{U}) \\ \mathcal{U} \cdot \mathcal{U}^T - \mu W \end{pmatrix} \qquad (28)$$

Therefore, writing the system Σ', equation (15), as

$$\frac{d\mathcal{Y}}{dt} = \gamma \begin{pmatrix} W_S \cdot S(\mathcal{V}) + I - S^{-1}\big(S(\mathcal{V})\big) \\ S(\mathcal{V}) \cdot S(\mathcal{V})^T - \tilde{\mu} W \end{pmatrix} + \gamma \begin{pmatrix} W_A . S(\mathcal{V}) \\ 0 \end{pmatrix},$$

where $\mathcal{Y} = (\mathcal{V}, W)^T$, we see that

$$\frac{d\mathcal{Y}}{dt} = -\gamma\Big(\nabla E\big(\sigma(\mathcal{V}, W)\big)\Big) + \Gamma(t) \tag{29}$$

where $\gamma(\mathcal{V}, W)^T = (\mathcal{V}, \epsilon W/M)^T$, $\sigma(\mathcal{V}, W) = (S(\mathcal{V}), W)$ and $\Gamma : \mathbb{R}_+ \to \mathcal{H}$ such that $\|\Gamma\| \underset{t\to+\infty}{\to} 0$ exponentially (because the system converges to $\mathcal{A}$). It follows that the time derivative of $\tilde{E} = E \circ \sigma$ along trajectories is given by:

$$\frac{d\tilde{E}}{dt} = \Big\langle \nabla \tilde{E}, \frac{d\mathcal{Y}}{dt} \Big\rangle = \Big\langle \nabla_{\mathcal{V}} \tilde{E}, \frac{d\mathcal{V}}{dt} \Big\rangle + \Big\langle \nabla_W \tilde{E}, \frac{dW}{dt} \Big\rangle. \tag{30}$$

Substituting equation (29) then yields

$$\begin{aligned} \frac{d\tilde{E}}{dt} &= -\Big\langle \nabla\tilde{E}, \gamma\big(\nabla E \circ \sigma\big) \Big\rangle + \underbrace{\Big\langle \nabla \tilde{E}, \Gamma(t) \Big\rangle}_{\tilde{\Gamma}(t)} \qquad (31) \\ &= -\Big\langle S'(\mathcal{V}) \nabla_{\mathcal{U}} E \circ \sigma, \nabla_{\mathcal{U}} E \circ \sigma \Big\rangle - \frac{\epsilon}{M} \Big\langle \nabla_W E \circ \sigma, \nabla_W E \circ \sigma \Big\rangle + \tilde{\Gamma}(t). \end{aligned}$$

We have used the chain–rule of differentiation, whereby

$$\nabla_V(\tilde{E}) = \nabla_V(E \circ \sigma) = S'(\mathcal{V}) \nabla_{\mathcal{U}} E \circ \sigma,$$

and $S'(\mathcal{V})\nabla_{\mathcal{U}} E$ (without dots) denotes the Hadamard (term by term) product, that is,

$$[S'(\mathcal{V})\nabla_{\mathcal{U}} E]_{ia} = s'(V_i^{(a)}) \frac{\partial E}{\partial U_i^{(a)}}$$

Note that $|\tilde{\Gamma}| \underset{t\to+\infty}{\to} 0$ exponentially because $\nabla \tilde{E}$ is bounded, and $S'(\mathcal{V}) > 0$ because the trajectories are bounded. Thus, there exists $t_1 \in \mathbb{R}_+$ such that $\forall t > t_1$, $\exists k \in \mathbb{R}_+^*$ such that

$$\frac{d\tilde{E}}{dt} \leq -k\|\nabla E \circ \sigma\|^2 \leq 0. \tag{32}$$

As in [Cohen and Grossberg, 1983] and [Dong and Hopfield, 1992], we apply the Krasovskii-LaSalle invariance principle [Khalil and Grizzle, 1996]. We check that:

- $\tilde{E}$ is lower bounded. Indeed, $\mathcal{V}$ and W are bounded. Given that $\mathcal{I}$ and S are also bounded it is clear that $\tilde{E}$ is bounded.
- $\frac{d\tilde{E}}{dt}$ is negative semidefinite on the trajectories as shown in equation (32).

Then the invariance principle tells us that the solutions of the system Σ' approach the set $M = \Big\{\mathcal{Y} \in \mathcal{H} : \frac{d\tilde{E}}{dt}(\mathcal{Y}) = 0\Big\}$. Equation (32) implies that $M = \Big\{Y \in \mathcal{H} : \nabla E \circ \sigma = 0\Big\}$. Since $\frac{d\mathcal{Y}}{dt} = -\gamma\Big(\nabla E \circ \sigma\Big)$ and $\gamma \neq 0$ everywhere, M consists of the equilibrium points of the system. This completes the proof.

8.3 Proof of theorem 4.2

Denote the right–hand side of system Σ', equation (15) by

$$F(\mathcal{V}, W) = \begin{cases} -\mathcal{V} + W \cdot S\big(\mathcal{V}\big) + I \\ \frac{\epsilon}{M}\big(S(\mathcal{V}).S(\mathcal{V})^T - \mu M W\big) \end{cases}$$

The fixed points satisfy the condition $F(\mathcal{V}, W) = 0$ which immediately leads to equations (19). Let us now check the linear stability of this system. The differential of F at $\mathcal{V}^*, W^*$ is

$$dF_{(\mathcal{V}^*,W^*)}(\mathcal{Z}, J) = \begin{pmatrix} -\mathcal{Z} + W^* \cdot \big(S'(\mathcal{V}^*)\mathcal{Z}\big) + J \cdot S(\mathcal{V}^*) \\ \frac{\epsilon}{M}\Big(\big(S'(\mathcal{V}^*)\mathcal{Z}\big) \cdot S(\mathcal{V}^*)^T + S(\mathcal{V}^*) \cdot \big(S'(\mathcal{V}^*)\mathcal{Z}\big)^T - \mu M J\Big), \end{pmatrix}$$

where $S'(\mathcal{V}^*)\mathcal{Z}$ denotes a Hadamard product, that is, $[S'(\mathcal{V}^*)\mathcal{Z}]_{ia} = s'(V_i^{*(a)})Z_i^{(a)}$. Assume that there exist $\lambda \in \mathbb{C}^*$, $(\mathcal{Z}, J) \in \mathcal{H}$ such that $dF_{(V^*,W^*)}\begin{pmatrix} \mathcal{Z} \\ J \end{pmatrix} = \lambda\begin{pmatrix} \mathcal{Z} \\ J \end{pmatrix}$. Taking the second component of this equation and computing the dot product with $S(\mathcal{V}^*)$ leads to

$$(\lambda + \epsilon\mu)J \cdot S = \frac{\epsilon}{M}\left((S'\mathcal{Z}) \cdot S^T \cdot S + S \cdot (S'\mathcal{Z})^T \cdot S\right)$$

where $S = S(\mathcal{V}^*)$, $S' = S'(\mathcal{V}^*)$. Substituting this expression in the first equation leads to

$$M(\lambda+\epsilon\mu)(\lambda+1)\mathcal{Z} = (\frac{\lambda}{\mu}+\epsilon)S\cdot S^T\cdot(S'\mathcal{Z})+\epsilon(S'\mathcal{Z})\cdot S^T\cdot S+\epsilon S\cdot(S'\mathcal{Z})^T\cdot S \tag{33}$$

Observe that setting $\epsilon = 0$ in the previous equation leads to an eigenvalue equation for the membrane potential only:

$$(\lambda + 1)\mathcal{Z} = \frac{1}{\mu M} S \cdot S^T \cdot (S'\mathcal{Z}).$$

Since $W^* = \frac{1}{\mu M}\big(S \cdot S^T\big)$, this equation implies that $\lambda + 1$ is an eigenvalue of the operator $X \mapsto W^*.(S'X)$. The magnitudes of the eigenvalues are always smaller than the norm of the operator. Therefore, we can say that if $1 > \|W^*\| S'_m$ then all the possible eigenvalues λ must have a negative real part. This sufficient condition for stability is the same as in [Faugeras et al., 2008]. It says that fixed points sufficiently close to the origin are always stable.

Let us now consider the case $\epsilon \neq 0$. Recall that $\mathcal{Z}$ is a matrix. We now "flatten" $\mathcal{Z}$ by storing its rows in a vector called $\mathcal{Z}_{row}$. We use the following result in [Brewer, 1978]: the matrix notation of operator $X \mapsto A \cdot X \cdot B$ is $A \otimes B^T$, where $\otimes$ is the Kronecker product. In this formalism the previous equation becomes

$$M(\lambda+\epsilon\mu)(\lambda+l)\mathcal{Z}_{row} = \bigg((\frac{\lambda}{\mu}+\epsilon) S \cdot S^T \otimes I_d + \epsilon I_d \otimes S^T \cdot S + \epsilon S \otimes S^T \bigg) \cdot (S'Z)_{\text{row}} \tag{34}$$

where we assume that the Kronecker product has the priority over the dot product. We focus on the linear operator $\mathcal{O}$ defined by the right hand side and bound its norm. Note that we use the following norm $\|W\|_\infty = \sup_X \frac{\|W.X\|}{\|X\|}$ which is equal to the largest magnitude of the eigenvalues of W.

$$\|\mathcal{O}\|_\infty \leq S'_m \bigg(|\frac{\lambda}{\mu}| \|S \cdot S^T \otimes I_d\|_\infty + \epsilon \|S \cdot S^T \otimes I_d\|_\infty + \epsilon \|I_d \otimes S^T \cdot S\|_\infty + \epsilon \|S \otimes S^T\|_\infty \bigg). \tag{35}$$

Define, ν_m to be the magnitude of the largest eigenvalue of $W^* = \frac{1}{\mu M}(S \cdot S^T)$. First, note that $S \cdot S^T$ and $S^T \cdot S$ have the same eigenvalues $(\mu M)\nu_i$ but different eigenvectors denoted by u_i for $S \cdot S^T$ and v_i for $S^T \cdot S$. In the basis set spanned by the $u_i \otimes v_j$, we find that $S \cdot S^T \otimes I_d$ and $I_d \otimes S^T \cdot S$ are diagonal with $(\mu M)\nu_i$ as eigenvalues. Therefore, $\|S \cdot S^T \otimes I_d\|_\infty = (\mu M)\nu_m$ and $\|I_d \otimes S^T \cdot S\|_\infty = (\mu M)\nu_m$. Moreover, observe that

$$(S^T \otimes S)^T \cdot (S^T \otimes S) \cdot (u_i \otimes v_j) = (S \cdot S^T \cdot u_i) \otimes (S^T \cdot S \cdot v_j) = (\mu M)^2 \nu_i \nu_j \, u_i \otimes v_j \tag{36}$$

Therefore, $(S^T \otimes S)^T \cdot (S^T \otimes S) = (\mu M)^2 \text{diag}(\nu_i \nu_j)$. In other words, $S^T \otimes S$ is the composition of an orthogonal operator (i.e. an isometry) and a diagonal matrix. Immediately, it follows that $\|S^T \otimes S\| \leq (\mu M)\nu_m$.

Compute the norm of equation (34)

$$|(\lambda + \epsilon\mu)(\lambda + 1)| \leq S'_m(|\lambda| + 3\epsilon\mu)\nu_m. \tag{37}$$

Define $f_\epsilon : \mathbb{C} \to \mathbb{R}$ such that $f_\epsilon(\lambda) = |(\lambda + \epsilon\mu)||(\lambda + 1)| - (|\lambda| + 3\epsilon\mu)S'_m\nu_m$. We want to find a condition such that $f_\epsilon(\mathbb{C}_+) > 0$, where $\mathbb{C}_+$ is the right half complex plane. This condition on $\epsilon,\ \mu,\ \nu_m,$ and S'_m will be a sufficient condition for linear stability. Indeed, under this condition we can show that only eigenvalues with a negative real part can meet the necessary condition (37). Complex number of the right half plane cannot be eigenvalues and thus the system is stable. The case $\epsilon = 0$ tells us that $f_0(\mathbb{C}_+) > 0$ if $1 > S'_m\nu_m$, compute

$$\frac{\partial f_\epsilon}{\partial \epsilon}(\lambda) = \mu(\Re(\lambda) + \mu\epsilon)\frac{|(\lambda + 1)|}{|(\lambda + \epsilon\mu)|} - 3\mu S'_m\nu_m$$

If $1 \geq \epsilon\mu$, which is most probably true given that $\epsilon << 1$, then $\frac{|(\lambda+1)|}{|(\lambda+\epsilon\mu)|} \geq 1$. Assuming $\lambda \in \mathbb{C}_+$ leads to:

$$\frac{\partial f_\epsilon}{\partial \epsilon}(\lambda) \geq \mu(\mu\epsilon - 3S'_m\nu_m) \geq \mu(1 - 3S'_m\nu_m)$$

Therefore, the condition $3S'_m\nu_m < 1$, which implies $S'_m\nu_m < 1$, and leads to $f_\epsilon(\mathbb{C}_+) > 0$.

8.4 A very short introduction to computational cohomology

In algebraic topology, topological spaces (which are continuous objects) can be classified by roughly counting their number of holes. This coordinate-invariant description of a topological space is called its homology (or cohomology, the difference between them is beyond the scope of this paper). In fact the homology can be summarized by giving the betti numbers of the topological state. The sequence of betti number is made of positive integers. The first three betti numbers have the following definition: the first is the number of connected components, the second is the number of two-dimensional or “circular” holes and the third is the number of 3-dimensional holes or “voids”. See [Hatcher, 2002] for a more rigorous approach.

However, in the example of toroidal retinotopy (see section 5.3.2 and figure 6). we are dealing with a discrete cloud of points. Therefore, one needs to extend the definition of the betti numbers to discrete objects in order to find the underlying topology of the space within which the points are distributed. This is called computational or persistent cohomology. One reconstructs the topological space by considering balls of a given radius centered on each point in the cloud. For each radius (the abscissa of the right picture of figure 6), one can compute the betti numbers of the resulting topological space. A barcode graph, e.g. the right picture of figure 6, is constructed by drawing a horizontal bar for each connected component, 2-dimensional hole or 3-dimenional void etc. This is done for a range of radii. Finding the persistent cohomology of a cloud of points consists in observing the set of betti numbers that are stable through a significantly large range of radii. One then assumes that the points most likely lie on a topological space of a given homology if the corresponding betti numbers are stable enough. See [Zomorodian and Carlsson, 2005] for details.

In section 5.3.2, we used the Jplex software package of [Sexton and Vejdemo-Johansson,] to compute the barcodes of the points corresponding to learning from inputs uniformly distributed over a 2-dimensional torus. We used 200 landmarks spread according to the maxminlandmark method to build simplices in Jplex, which returned the maximum radius (beyond which all the betti numbers except the first vanish) we used in the simulation. In figure 6, we see that for a wide range of radii the triplet $(1, 2, 1)$ is stable. This corresponds to a 2 dimensional torus.

References

[Amari, 1998] Amari, S. (1998). Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276.

[Amari et al., 1992] Amari, S., Kurata, K., and Nagaoka, H. (1992). Information geometry of Boltzmann machines. *IEEE Transactions on Neural Networks*, 3(2):260–271.

[Bartsch and Van Hemmen, 2001] Bartsch, A. and Van Hemmen, J. (2001). Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex. *Biological Cybernetics*, 84(1):41–55.

[Bi and Poo, 2001] Bi, G. and Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual review of neuroscience*, 24:139.

[Bienenstock et al., 1982] Bienenstock, E., Cooper, L., and Munro, P. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci*, 2:32–48.

[Borg and Groenen, 2005] Borg, I. and Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications*. Springer Verlag.

[Bosking et al., 1997] Bosking, W., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, 17(6):2112.

[Bressloff, 2005] Bressloff, P. (2005). Spontaneous symmetry breaking in self–organizing neural fields. *Biological Cybernetics*, 93(4):256–274.

[Bressloff et al., 2001] Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., and Wiener, M. (2001). Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Phil. Trans. R. Soc. Lond. B*, 306(1407):299–330.

[Bressloff and Cowan, 2003] Bressloff, P. C. and Cowan, J. D. (2003). A spherical model for orientation and spatial frequency tuning in a cortical hypercolumn. *Philosophical Transactions of the Royal Society B*.

[Brewer, 1978] Brewer, J. (1978). Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25(9).

[Chklovskii et al., 2002] Chklovskii, D., Schikorski, T., and Stevens, C. (2002). Wiring optimization in cortical circuits. *Neuron*, 34(3):341–347.

[Chossat and Faugeras, 2009] Chossat, P. and Faugeras, O. (2009). Hyperbolic planforms in relation to visual edges and textures perception. *PLoS Computational Biology*, 5(12):367–375.

[Cohen and Grossberg, 1983] Cohen, M. and Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. In *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, pages 815–826.

[Coifman et al., 2005] Coifman, R., Maggioni, M., Zucker, S., and Kevrekidis, I. (2005). Geometric diffusions for the analysis of data from sensor networks. *Current opinion in neurobiology*, 15(5):576–584.

[Coombes, 2005] Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biological Cybernetics*, 93(2):91–108.

[Cottet, 1995] Cottet, G. (1995). Neural networks: Continuous approach and applications to image processing. *Journal of Biological Systems*, 3:1131–1139.

[Dayan and Abbott, 2001] Dayan, P. and Abbott, L. (2001). *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press.

[Degond and Mas-Gallic, 1989] Degond, P. and Mas-Gallic, S. (1989). The Weighted Particle Method for Convection-Diffusion Equations. Part 1: The Case of an Isotropic Viscosity. *Mathematics of Computation*, pages 485–507.

[Dong and Hopfield, 1992] Dong, D. and Hopfield, J. (1992). Dynamic properties of neural networks with adapting synapses. *Network: Computation in Neural Systems*, 3(3):267–283.

[Edwards, 1996] Edwards, R. (1996). Approximation of neural network dynamics by reaction-diffusion equations. *Mathematical methods in the applied sciences*, 19(8):651–677.

[Faugeras et al., 2008] Faugeras, O., Grimbert, F., and Slotine, J.-J. (2008). Abolute stability and complete synchronization in a class of neural fields models. *SIAM J. Appl. Math*, 61(1):205–250.

[Földiák, 1991] Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3(2):194–200.

[Geman, 1979] Geman, S. (1979). Some averaging and stability results for random differential equations. *SIAM J. Appl. Math*, 36(1):86–105.

[Gerstner and Kistler, 2002] Gerstner, W. and Kistler, W. M. (2002). Mathematical formulations of hebbian learning. *Biological Cybernetics*, 87:404–415.

[Hatcher, 2002] Hatcher, A. (2002). *Algebraic topology*. Cambridge Univ Pr.

[Hebb, 1949] Hebb, D. (1949). *The organization of behavior: a neuropsychological theory.* Wiley, NY.

[Hubel and Wiesel, 1977] Hubel, D. H. and Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proc. Roy. Soc. B*, 198:1–59.

[Khalil and Grizzle, 1996] Khalil, H. and Grizzle, J. (1996). *Nonlinear systems*. Prentice hall Upper Saddle River, NJ.

[Kohonen, 1990] Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE*, 78(9).

[Lawlor and Zucker, 2010] Lawlor, M. and Zucker, S. (2010). Third-Order Edge Statistics Reveal Curvature Dependency. In *Snowbird workshop on learning*.

[Miikkulainen et al., 2005] Miikkulainen, R., Bednar, J., Choe, Y., and Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. Springer, New York.

[Miller, 1996] Miller, K. (1996). Synaptic economics: competition and cooperation in synaptic plasticity. *Neuron*, 17:371–374.

[Miller and MacKay, 1996] Miller, K. and MacKay, D. (1996). The role of constraints in hebbian learning. *Neural Comp*, 6:100–126.

[Miller et al., 1989] Miller, K. D., Keller, J. B., and Stryker, M. P. (1989). Ocular dominance column development: analysis and simulation. *Science*, 245:605–615.

[Oja, 1982] Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biology*, 15:267–273.

[Ooyen, 2001] Ooyen, A. (2001). Competition in the development of nerve connections: a review of models. *Network: Computation in Neural Systems*, 12(1):1–47.

[Petitot, 2003] Petitot, J. (2003). The neurogeometry of pinwheels as a sub-Riemannian contact structure. *Journal of Physiology-Paris*, 97(2-3):265–309.

[Sejnowski and Tesauro, 1989] Sejnowski, T. and Tesauro, G. (1989). The Hebb rule for synaptic plasticity: algorithms and implementations. *Neural models of plasticity: Experimental and theoretical approaches*, pages 94–103.

[Sexton and Vejdemo-Johansson,] Sexton, H. and Vejdemo-Johansson, M. JPlex simplicial complex library. `http://comptop.stanford.edu/programs/jplex/`.

[Swindale, 1996] Swindale, N. (1996). The development of topography in the visual cortex: a review of models. *Network: Computation in Neural Systems*, 7(2):161–247.

[Takeuchi and Amari, 1979] Takeuchi, A. and Amari, S. (1979). Formation of topographic maps and columnar microstructures in nerve fields. *Biological Cybernetics*, 35(2):63–72.

[Tikhonov, 1952] Tikhonov, A. (1952). Systems of differential equations with small parameters multiplying the derivatives. *Matem. sb*, 31(3):575–586.

[Verhulst, 2007] Verhulst, F. (2007). Singular perturbation methods for slow–fast dynamics. *Nonlinear Dynamics*, 50(4):747–753.

[Wallis and Baddeley, 1997] Wallis, G. and Baddeley, R. (1997). Optimal, unsupervised learning in invariant object recognition. *Neural computation*, 9(4):883–894.

[Zomorodian and Carlsson, 2005] Zomorodian, A. and Carlsson, G. (2005). Computing persistent homology. *Discrete and Computational Geometry*, 33(2):249–274.